

Research Data Management - Doctoral Student

Basic information

1. Name of the interviewee:

2. Department or center:

3. Institution:

This worksheet is designed to elicit the information necessary to develop a data curation profile about data from a particular research project. In your responses to this worksheet and to the interview

questions, please limit your focus to the data associated with the particular research project you have selected.

This worksheet is meant to be filled out as a part of the interview and your responses to the questions in each module will serve to guide the conversation. The interviewer will then ask you some follow up questions about your responses to gather additional details and to better understand your priorities and needs.

If you have any questions, need more information, or would like clarification about any item listed in this worksheet, please do not hesitate to ask your interviewer.

Information on the research project and the interview

Participation is voluntary.

Information gathered by interviews will be used for planning research data management (RDM) trainings and services for doctoral students.

The interview e-form is part of the interview and works as an introduction to different phases of the RDM. Interviewers can ask some additional questions besides of the e-form.

Interviewees can interrupt their participation in the research project or refuse to answer to the questions any time.

Endurance of the interview is about 1-2 hours. Interviewee fills the e-form during the interview one module at a time.

Confidentiality in the research publication and in the research data. Preservation and potential reuse of the research data

The confidentiality of the interview will be secured so that any materials that include personal information will not be handed over to any third party.

Names of the interviewees or names of other people that possibly come out during the interview will be deleted or changed for code names in the research publication and in the archived research data.

The status (doctoral student), organization, faculty, department, discipline and sex will be told.

Research publication can include direct citations with no names of the interviewees, but with the organization, status, faculty, department, discipline and sex of the interviewee.

When the research project has ended, the excel sheet containing interviewees' answers to the questions, will be handed over the Data Archive to be permanently preserved and used for research, education and studying purposes.

More identifiers will be deleted during the archiving process (anonymisation) if needed.

Information on the Data Archive: http://www.fsd.uta.fi/shared/resources/2014_esite_tutkittaville_ENG.pdf

Interviewers will gladly answer all the questions concerning this interview.

4. Luottamuksellisuuden turvaaminen ja tutkimusaineiston säilyttäminen yllä kuvatulla tavalla / Approval of securing the confidentiality and preserving the research data as told above

- Hyväksyn yllä kerrotun. / I approve
- Hyväksyn yllä kerrotun seuraavin varauksin / I approve with next reservations

Module 1 - The Data Sets

5. Please provide a brief description of the data:

Module 2 - The Lifecycle of the Data Set

Initial Data Stage

6. Initial Data Stage:

7. Initial Data Stage: Approximately how many data files exist at this stage?

8. Initial Data Stage: What is the approximate average size of each file at this stage? (Please include the unit of measurement - kb, MB, GB, etc.)

9. Initial Data Stage: What format(s) are the data in? (For example MS-Excel 2007, MySQL database, JPEG 2000 images, a raw data file form a Chambell CR10 data logger, etc.)

Second Data Stage

10. Second Data Stage:

11. Second Data Stage: Approximately how many data files exist at this stage?

12. Second Data Stage: What is the approximate average size of each file at this stage? (Please include the unit of measurement - kb, MB, GB, etc.)

13. Second Data Stage: What format(s) are the data in? (E.g. MS-Excel 2007, MySQL database, JPEG 2000 images, a raw data file form a Chambell CR10 data logger, etc.)

Third Data Stage

14. Third Data Stage:

15. Third Data Stage: Approximately how many data files exist at this stage?

16. Third Data Stage: What is the approximate average size of each file at this stage? (Please include the unit of measurement - kb, MB, GB, etc.)

17. Third Data Stage: What format(s) are the data in? (E.g. MS-Excel 2007, MySQL database, JPEG 2000 images, a raw data file from a Chambell CR10 data logger, etc.)

Fourth Data Stage

18. Fourth Data Stage:

19. Fourth Data Stage: Approximately how many data file exist at this stage?

20. Fourth Data Stage: What is the approximate average size of each file at this stage? (Please include the unit of measurement - kb, MB, GB, etc.)

21. Fourth Data Stage: What format(s) are the data in? (E.g. MS-Excel 2007, MySQL database, JPEG 2000 images, a raw data file form a Chambell CR10 data logger, etc.)

Fifth Data Stage

22. Fifth Data Stage:

23. Fifth Data Stage: Approximately how many data files exist at this stage?

24. Fifth Data Stage: What is the approximate average size of each file at this stage? (Please include the unit of measurement - kb, MB, GB, etc.)

25. Fifth Data Stage: What format(s) are the data in? (E.g. MS-Excel 2007, MySQL database, JPEG 2000 images, a raw data file form a Chambell CR10 data logger, etc.)

Additional question:

26. Did you inherit any part of the data set that you have described from another student or researcher?

- Yes
- No
- I don't know

Module 3 – Data Management

27. How and where do you store your data? (If more than one storage method is used, please list all of them).

28. Are there formal policies or procedures that you follow in managing your data set?

- Yes
- No
- I don't know

29. Do you do any of the following as a part of managing your data set?

	Always	Sometimes	Occasionally	Never	I don't know
Use a file naming system to assign names to your data files?	<input type="radio"/>				
Make back-up copies of your data?	<input type="radio"/>				
Take security measures to protect your data?	<input type="radio"/>				
Follow a system to identify and track different versions of your data files as they progress through the data lifecycle?	<input type="radio"/>				
Provide a means to identify the "official" or "authoritative" version of your data files?	<input type="radio"/>				

Module 4- Tools

30.

What tools, including software or hardware, are used in generating your data? (e.g., "a data logger", "a remote sensor", etc. – please describe.)

31.

What tools, including software or hardware, are required to analyze or utilize the data? (e.g., "Microsoft Excel 2007", "ArcGIS", "I wrote my own program", etc. please describe.)

32.

Do you generate any visual representations of your data? (Visual representations include charts, graphs, plots, videos, images, etc.)

- Yes
- No
- I don't know

33.

Over the course of your research, have you or do you anticipate having to convert any part of your data from one format into another? (For example, converting a Microsoft 2007 file into a .csv file)

- Yes
- No
- I don't know

Module 5 - Documentation and Description of Data

34.

Please explain briefly how your datasets are documented and how they have been described (e.g. "detailed annotations", "a lab notebook", "a code book", "a data dictionary", "column headings in a spreadsheet", etc.).

If you have used any standardized forms of documentation, description or metadata please identify the standard(s) in your response.

35.

Is the amount of documentation and description that you currently create sufficient for another person with similar expertise to be able reproduce your data set on his or her own?

- Yes
- No

I don't know

36.

Is the amount of documentation and description that you currently create sufficient for another person with similar expertise to be able to understand and use the data for their own purposes?

- Yes
- No
- I don't know

37.

Are there standard practices in your discipline for organizing, documenting or describing data that you are aware of?

- Yes
- No
- I don't know

38. Are you familiar with the concept of metadata?

- Yes
- No
- I don't know

Module 6 – Acquiring and Using External Data

39.

Have you ever made use of data that was generated by a source outside of your project? This could include downloading data from a national registry, data repository, accessing data from a supplemental file, requesting a data set from another researcher, etc.

- Yes
- No
- I don't know

40. Do you know how to cite a data set?

- Yes
- No
- I don't know

41. What are the major data repositories or sources of data in your discipline (if any)?

Module 7: Sharing data

42. Have you ever deposited any data into a data repository?

- Yes
- No
- I don't know

43. Are you planning to deposit your data into a data repository in the future?

- Yes
- No
- I don't know

44. Would you be willing to share your data with someone outside of your project?

- Yes
- No
- I don't know

45.

If you answered “yes” to Question 49, would you place any conditions on sharing this data with someone outside of your project?

- Yes
- No
- I don't know

46.

In the journals of your field, are data or other supplemental information, accepted for publication?

- Yes
- No
- I don't know

47.

Who would you imagine might be interested in the dataset that you are producing? (For example, other researchers in my field, researchers outside of my field, practicing professionals, policy makers, etc.)

48.

How would you imagine this data being used by the groups / people you listed in the previous question? What value would the data have for these groups / people?

49. Will you leave a copy of your data with your advisor and/or project when you graduate?

- Yes

- No
- I don't know

Module 8 – Intellectual Property & Ethics

50. Who is the owner of the data?

51. Is this dataset bound by any privacy or confidentiality concerns?

- Yes
- No
- I don't know

52.

Do you feel that you have a good understanding of the university's policies on research data?

- Yes
- No
- I don't know

53.

Have you received any training or instruction on intellectual property issues as they relate to data and/or data ethics generally?

- Yes
- No
- I don't know

Module 9 – Data Preservation

54.

What are the most important parts of your data set to preserve (manage and maintain over time)?

55. How long would your data set be useful or have value for you or others if it were to be preserved?

- My dataset does not need to be preserved.
- Less than 3 years.
- 3 years or more but less than 5 years.
- 5 years or more but less than 10 years.
- 10 years or more but less than 20 years.
- 20 years or more but less than 50 years.
- 50 years or more but less than 100 years.
- Indefinitely.
- I don't know

56. Are you or anyone else in your project taking steps to preserve this data set?

- Yes
- No
- I don't know

Module 10 – Educational Needs

Please indicate how important you believe it is for you to be knowledgeable in each of the competencies listed below by the time you graduate. Please also tell how well do you think you'll manage the competence now.

Discovery and Acquisition of Data

Skills may include:

Locates and utilizes disciplinary data repositories or other external data sources. Evaluates the quality of the data available from external sources. Not only identifies appropriate external data sources, but also imports data and converts it when necessary, so it can be used locally.

57.

Discovery and Acquisition of Data: Importance

- Not important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

58. Discovery and Acquisition of Data: Competence

- Don't have competence
- Somewhat competence
- Good competence
- Very good competence
- Ultimate competence

Databases and Data Formats

Skills may include:

Understands the concept of relational databases, how to query those databases, and becomes familiar with standard data formats and types for their discipline. Understands which formats and data types are appropriate for different research questions.

59.

Databases and Data Formats: Importance

- Not important

- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

60. Databases and Data Formats: Competence

- Don't have
- Somewhat
- Good
- Very good
- Ultimate

Data Conversion and Interoperability

Skills may include:

Is proficient in migrating data from one format to another. Understands the risks and potential loss or corruption of information caused by changing data formats. Understands the benefits of making data available in standard formats to facilitate downstream use.

61.

Data Conversion and Interoperability: Importance

- Not important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

62. Data Conversion and Interoperability: Competence

- Don't have

- Somewhat
- Good
- Very good
- Ultimate

Data Processing and Analysis

Skills may include:

Is familiar with the basic data processing and analysis tools and techniques of the discipline or research area. Understands the effect that these tools may have on the data. Uses appropriate workflow management tools to automate repetitive analysis of data.

63.

Data Processing and Analysis: Importance

- Not Important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

64. Data Processing and Analysis: Competence

- Don't have
- Somewhat
- Good
- Very good
- Ultimate

Data Visualization and/or Representation

Skills may include:

Proficiently uses basic visualization tools of discipline. Avoids misleading or ambiguous representations when presenting data in tables, charts, diagrams, etc. Chooses the appropriate type of visualization, such as maps, graphs, animations, or videos, based on their understanding of the reason / purpose for

visualizing or displaying data.

65.

Data Visualization and/or Representation: Importance

- Not Important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

66. Data Visualization and/or Representation: Competence

- Don't have
- Somewhat
- Good
- Very good
- Ultimate

Data Management and Organization

Skills may include:

Understands the lifecycle of data, develops data management plans, and keeps track of the relation of subsets or processed data to the original data sets. Creates standard operating procedures for data management and documentation.

67.

Data Management and Organization: Importance

- Not Important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

68. Data Management and Organization: Competence

- Don't have
- Somewhat
- Good
- Very good
- Ultimate

Data Quality and Documentation

Skills may include:

Recognizes, documents, and resolves any apparent artifacts, incompleteness, or corruption of data sets. Utilizes metadata to facilitate an understanding of potential problems with data sets. Documents data sufficiently enough to enable the reproduction of the research results and the data by others. Tracks data provenance and clearly delineates and denotes versions of a data set.

69.

Data Quality and Documentation: Importance

- Not Important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

70. Data Quality and Documentation: Competence

- Don't have
- Somewhat
- Good
- Very good
- Ultimate

Metadata and Data Description

Skills may include:

Understands the rationale for metadata and proficiently annotates and describes data so it can be understood and used by self and others. Develops the ability to read and interpret metadata from external disciplinary sources. Understands the structure and purpose of ontologies in facilitating better sharing of data.

71.

Metadata and Data Description: Importance

- Not Important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

72. Metadata and Data Description: Competence

- Don't have
- Somewhat
- Good
- Very good
- Ultimate

Cultures of Practice

Skills may include:

Recognizes the practices, values, and norms of his/her chosen field, discipline, or subdiscipline as they relate to managing, sharing, curating, and preserving data. Recognizes relevant data standards of his/her field (metadata, quality, formatting, etc.) and understands how these standards are applied.

73.

Cultures of Practice: Importance

- Not Important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

74. Cultures of Practice: Competence

- Don't have
- Somewhat
- Good
- Very good
- Ultimate

Ethics and Attribution

Skills may include:

Develops an understanding of intellectual property, privacy and confidentiality issues, and the ethos of the discipline when it comes to sharing and administering data. Acknowledges data from external sources appropriately. Avoids misleading or ambiguous representations when presenting data.

75.

Ethics and Attribution: Importance

- Not Important
- Somewhat Important
- Important

- Very Important
- Essential
- I don't know or NA

76. Ethics and Attribution: Competence

- Don't have
- Somewhat
- Good
- Very good
- Ultimate

Data Curation and Re-use

Skills may include:

Recognizes that data may have value beyond the original purpose, to validate research, or for use by others. Is able to distinguish which elements of a data set are likely to have future value for self and for others. Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community-driven e-research. Recognizes that data must be prepared for its eventual curation at its creation and throughout its lifecycle. Articulates the planning and activities needed to enable data curation, both generally and within his/her local practice. Understands how to cite data as well as how to make his/her data citable.

77.

Data Curation and Re-use: Importance

- Not Important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

78. Data Curation and Re-use: Competence

- Don't have

- Somewhat
- Good
- Very good
- Ultimate

Data Preservation

Skills may include:

Recognizes the benefits and costs of data preservation. Understands the technology, resources, and organizational components of preserving data. Utilizes best practices in preparing data for its eventual preservation during its active lifecycle. Articulates the potential long term value of his/her data for him/herself or others and is able to determine an appropriate preservation timeframe. Understands the need to develop preservation policies and is able to identify the core elements of such policies.

79.

Data Preservation: Importance

- Not Important
- Somewhat Important
- Important
- Very Important
- Essential
- I don't know or NA

80. Data Preservation: Competence

- Don't have
- Somewhat
- Good
- Very good
- Ultimate